DOI 10.18699/SBB-2022-347

The new method of DNA barcoding

Kiryanova O.[1]*, Gubaydullin I.[1, 2], Chemeris A.[3]

[1] *Ufa State Petroleum Technological University, Ufa, Russia*

[2] *Institute of Petrochemistry and Catalysis – Subdivision of the Ufa Federal Research Center, RAS, Ufa, Russia*

[3] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Center, RAS, Ufa, Russia*

** olga.kiryanova27@gmail.com*

Key words: DNA barcoding, DNA identification, *in silico* genome analysis

Motivation and Aim: The classification and systematization of living organisms is currently an actual task. A popular method for genomes classification is DNA barcoding. This method requires the presence of a reference DNA fragment, which allows to explicitly identify living organisms [1]. The cytochrome oxidase gene was assumed as a reference region for the living organisms. However, this gene does not lead to the satisfactory results for the plants systematization. Here we propose a new method for DNA certification of living organisms on the example of plants.

Methods and Algorithms: We have utilized *in silico* multiplex PCR analysis. Namely, we have found the assumed primer annealing sites in a genome. Search of the annealing positions was performed using the Boyer-Moore algorithm, where the primers are samples and genome is a text. The amplicons corresponding to the annealing sites were converted into a binary format. It should be noted, that amplicons were selected in the range from 51 to 500 nucleotides. Such amplicons have been unambiguously determined experimentally to within a nucleotide. The binary sequence is easy enough to translate into a barcode format. An example of the resulting barcode is presented on Fig. 1.

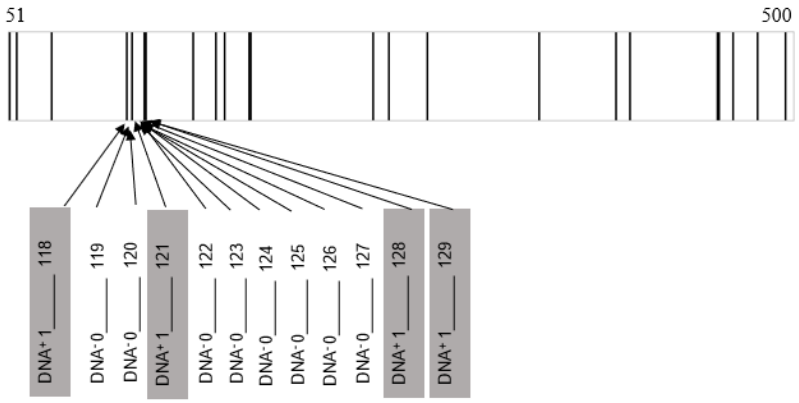

Fig. 1. The principle of constructing a barcode

Results: We have conducted a comparative analysis of the following plant genomes: wheat (6 genomes), hevea (3 genomes), and arabidopsis (9 genomes). The proposed approach was utilized to obtain barcodes for each group. Each set of barcodes indicated the differences and similarities for certain plants. It should be pointed out, that different primers provide different barcodes for the same genome. In this case, computer analysis

makes it possible to identify successful primers for the experimental PCR. As a result, the optimal number of primers could be determined for DNA certification. However, primers and amplicons should be treated as a whole to ensure the uniqueness of any DNA barcode and provide a successful distinction between living organisms.

Conclusion: The presented DNA certification can reveal similarities and differences at the level of large order taxa (species, family) however it more reliable for the certification of cultivars. Hence, it allows to test several primers as well as to get an idea of the success experiment. Thus, the proposed method provides a reliable and convenient way to store genetic data in a digital format, and reduce the time consumption in PCR tests planning.

Acknowledgements: The study was funded by Russian Foundation for Basic Research (17-44-020120).

References

1. Shneyer V.S. DNA barcoding of animal and plant species as an approach for their molecular identification and describing of diversity. *J General Biol*. 2009;70(4):296-315.
2. Gasfield D. Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. St. Petersburg: Nevskii dialect, 2003.